# DOCUMENT RESUME

ED 092 581 TM 003 692

AUTHOR Neel, John H.; Stallings, William M.
TITLE A Monte Carlo Study of Levene's Test of Homogeneity of Variance: Empirical Frequencies of Type I Error in Normal Distributions.
PUB DATE [Apr 74]
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association Convention (Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Analysis of Variance; Educational Research; *Hypothesis Testing; Sampling; *Statistical Analysis; *Statistical Bias; Tests of Significance
IDENTIFIERS *Monte Carlo Methods

ABSTRACT

An influential statistics test recommends a Levene text for homogeneity of variance. A recent note suggests that Levene's test is upwardly biased for small samples. Another report shows inflated Alpha estimates and low power. Neither study utilized more than two sample sizes. This Monte Carlo study involved sampling from a normal population for all combinations of two to seven variances and equal sample sizes from three to twelve. Alpha was upwardly biased for smaller sample sizes; Alpha was upwardly biased for increasing numbers of variances and fixed sample sizes. The Levene technique apparently has little to recommend it besides computational simplicity. (Author)

ED092581



# A MONTE CARLO STUDY OF LEVENE'S TEST OF HOMOGENEITY OF VARIANCE: EMPIRICAL FREQUENCIES OF TYPE I ERROR IN NORMAL DISTRIBUTIONS

John H. Neel and William M. Stallings
Georgia State University

TM 003 692

Presented at
the American Educational Research Association Convention
Chicago, April 18, 1974

# ABSTRACT

An influential statistics text recommends a Levene text for homogeneity of variance. A recent note suggests that Levene's test is upwardly biased for small samples. Another report shows inflated $\alpha$ estimates and low power. Neither study utilized more than two sample sizes. This Monte Carlo study involved sampling from a normal population for all combinations of two to seven variances and equal sample sizes from three to twelve. Alpha was upwardly biased for smaller sample sizes; $\alpha$ was upwardly biased for increasing numbers of variances and fixed sample sizes. The Levene technique apparently has little to recommend it besides computational simplicity.

A recently published and highly regarded educational statistics text (Glass and Stanley, 1970) recommends, with little qualification, the use of a Levene test (1960) for homogeneity of variance. Orginally the Levene test was popularized for educational researchers by Glass (1966). Briefly, the particular Levene test advocated by Glass and Stanley (1970) and Glass (1966) involves an analysis of variance using as data the absolute values of the deviation scores, where the deviation scores are deviations from the column means (Levene recommended several other versions of this test, all involving an analysis of variance on a monotorically increasing function of the deviation scores).

Although Glass and Stanley (1970) do state that the robustness of the Levene technique has been demonstrated only "for equal numbers of observations in J samples", they nevertheless give the impression that the test is a statistically desirable substitute for more standard tests such as Bartlett's, Scheffe's, Cochran's, and Hartley's. Is this impression correct?

In the original exposition, Levene (1960) reported the results of Monte Carlo studies based upon several combinations of equal sized samples ($n$) and number of groups ($J$). When the sample size in each level was relatively large ($n = 10$ and $n = 20$) and the number of levels small ($J = 2$ and $J = 3$), the empirical probabilities of Type 1 error were quite close to the nominal values of $\alpha$. Even with Levene's data, however, when $J = 10$ and $n = 20$, the empirical probability was .063 at the nominal .05 level.

After what some might consider an inordinate lag, several theoretical and empirical studies on the Levene technique have been published. In a letter to the editor of Technometrics, Miller (1972) reported a small scale Monte Carlo study of the Levene test for three different distributions. In sampling from a normal distribution with $J = 2$ (apparently) and $n = 3$ and $n = 4$, Miller found inflated ratios of $E(S_1^2)$ to $E(S_2^2)$. Miller's implicit conclusion was that Levene's test probably was inappropriate for small $n$'s.

Games et al. (1972) reported a comparative study of several homogeneity of variance tests, also using Monte Carlo techniques. For one combination ($J = 3$, $n = 6$) and a variety of distributions, Games et al. (1972) found inflated empirical alpha values for the Levene test. Moreover, the power of the Levene test was low. Incidentally, both Games et al. (1972) and Miller (1972) seem to attribute part of the problem with the Levene test to the dependence of observations within each column.

Miller (1968) demonstrated that the Levene test is not asymptotically distribution free, that is, things don't necessarily get better as n increases.

Taken together, the studies by Miller (1968,1972), and Games et al. (1972) still do not constitute a needed overkill of the Levene technique. First, this failure stems from a paucity of investigations, concerning the technique, over a wide range of sample sizes and number of variances when a normal distribution is assumed. If, indeed, the statistical properties are inadequate for normal distributions, investigations using non-normal distributions are merely of academic interest. Second, this failure results from none of the reviewed studies employing more than two combinations of sample sizes and number of variances tested. Thus, any trends in power or significance level could not be detected by these studies.

This note summarizes the results of a fairly extensive Monte Carlo study. (All programming was done by the senior author). The extent of this study involved examining the Levene technique for all combinations of two to seven

TABLE 1

Observed Number of Type 1 Errors In
10,000 Tests, Given a Nominal
Level of .05

| nber of Variances (J) | Sample Size (n ) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2 | 525 | 791 | 746 | 659 | 640 | 601 | 614 | 576 | 562 | 549 |
| 3 | 711 | 879 | 774 | 737 | 669 | 604 | 624 | 518 | 556 | 524 |
| 4 | 1138 | 959 | 788 | 781 | 725 | 597 | 609 | 583 | 566 | 516 |
| 5 | 1243 | 1135 | 903 | 710 | 723 | 610 | 648 | 614 | 570 | 507 |
| 6 | 1598 | 1174 | 934 | 902 | 747 | 638 | 634 | 613 | 717 | 66[illegible] |
| 7 | 1843 | 1443 | 1140 | 1128 | 867 | 670 | 798 | 721 | 674 | 539 |

variances and equal sample size of three through twelve in a one-way ANOVA framework. At each combination, 10,000 sample *F*'s were creating using a normal population. To create each sample, pseudo-random numbers were generated using the multiplicative congruential method, and these were then transformed to standard normal variates (For a detailed discussion of the technique see Naylor *et al.*, 1966, Pp. 51 & 95). These standard normal variates were then placed in an analysis of variance framework. For example, with *J* = 3 and *n* = 5, fifteen standard normal deviates were generated and placed in three groups of 5 each. The Levene test was then applied to test the homogeneity of variance of the three groups or levels. The resulting *F* was one of the 10,000 *F*'s generated at this combination of *J*=3 and *n* = 5.

Since 10,000 samples were selected in each case, the expected number of *F*'s significant at the .05 level is 500. The obtained number of *F*'s significant at the .05 level is reported in Table 1.

In *all* instances the Levene technique produced upwardly biased $\alpha$ estimates. Two interacting trends seem discernable in this upward bias: (a) For smaller *n* (roughly, *n* less than 9), the bias increases as the number of variances tested increases; (b) for larger *n*, the bias seems to remain constant or even to fall as the number of variances increases. As would be expected in a sampling study, there are a few minor anomalies in the data but none which essentially detracts from our conclusions. Overall, the results of our study bear out the theoretical speculations of Miller (1968) and support the Monte Carlo studies of Miller (1972) and of Games *et al.* (1972). Not only have we supported their studies, but by covering a wide range of sample sizes and number of variances we have provided additional data on the upwardly biased $\alpha$ level of the statistic. Since the Levene technique seems to have little to offer the educational researcher except computational simplicity, its use should be discouraged.

# REFERENCES

Games, P. A., Winkler, H. B., & Probert, D. A. Robust tests for homogeneity of variance. Educational and Psychological Measurement, 1972, 32 (4), 887-910.

Glass, G. V Testing homogeneity of variances. American Educational Research Journal, 1966, 3 (3), 187 - 190.

Glass, G. V & Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Leven, H. Robust test for equality of variances. In I. Olkin (ed.), Contributions to probability and statistics. Stanford, California: Stanford University Press, 1960, 278 - 292.

Miller, A. J. Letter to the editor. Technometrics, 1972, 14 (2), 507.

Miller, R. G. Jackknifing variances. Annals of Mathematical Statistics, 1968, 39, 567 - 582.

Naylor, T. H.; Balintfy, T. L.; Burdick, D. S.; and Chu, Kong Computer simulation techniques. New York: Wiley, 1972.